Title: Probabilistic Factoring

Brief Overview:

Every integer is a product of powers of prime numbers. We are interested in the probability that a relatively large positive random integer has a factor in a random set of relatively small primes which is determined by the flipping of a fair virtual coin.

In this simulation activity, students are given the opportunity to investigate and apply varied mathematical concepts through the construction of their own solution algorithm. The activity and related mathematics are not computer language dependent. The emphasis has been placed on problem solving and the related mathematics rather than on a specific programming language.

Links to NCTM Standards:

- **Mathematics as Problem Solving**
  One of the most significant applications of computers is simulation. Methods implemented in programming algorithms allow individuals to investigate patterns of a process without actually conducting the experimental procedure. Furthermore, programming algorithms can serve as key tools in investigating some applications which are otherwise inaccessible to many.

- **Mathematics as Communication**
  Students must propose and justify their solution strategies for estimating the related theoretical probability to their peers in a brainstorming session. In addition, they may be required to submit a written presentation of their procedure and/or the class procedure.

- **Mathematics as Reasoning**
  Students must determine how the mathematical properties and procedures related to the problem under consideration may be implemented in a simulation algorithm.

- **Mathematical Connections**
  Students will use concepts of mathematics in a modeling/simulation process.

- **Probability**
  Students will apply relative frequency probability techniques to approximate the value of a theoretical probability.

- **Discrete Mathematics**
  Students will apply counting procedures and number theory concepts in the development of a simulation algorithm.

Grade/Level:

11-12

Duration/Length:

1 hour to present and discuss the activity. Completion time of the related program simulation will vary with the background of the students.

Prerequisite Knowledge:

Students should have working knowledge of the following skills:

- Factorization of integers into products of powers of prime numbers
- Familiarity with logarithms and their properties
- Familiarity with a random number generator in the computer language of the student's choice

Objectives:

- Students will simulate the game described below to estimate the probability of winning and compare their estimate to a theoretical estimate.
- Students will learn about the prime number theorem and use it to estimate the probability of losing the game due to a particular circumstance, which in turn gives an estimate of the best case probability of winning.

Materials/Resources/Printed Materials:

- A computer with word size of at least 32 bits is best. This includes most machines that are less than 10 years old. For smaller sizes one should adjust the parameters M and e described below so that M+e is as big as it can be without being larger than word size.

Development/Procedures:

The problem is best described in terms of the following game.

Let P[1],P[2], ... ,P[n] denote the first n prime numbers. An integer N, which is much larger than P[n] is chosen at random. For each prime P[i] (1<= i <= n) a fair virtual coin is flipped. If the coin comes up tails then we proceed to the next prime P[i+1] (if i < n), and if the coin comes up heads, N is checked for divisibility by P[i]. If P[i] divides N, then a factor of N has been found and the game is won, otherwise we proceed to the next prime P[i+1] (if i< n). The game is lost if no factor of N has been found after n coin flips.

Note that if N itself is prime, then the game will be lost regardless of what happens on the coin tosses, since we have required that N is greater than P[n].

Procedures:

Write a program to simulate this game being played several times. Some suggestions for parameters follow. Start with n=100, i.e., consider the first 100 primes. They can be found once and then stored in an array . Choose N in the interval [M-e,M+e], where M= 2^23 and e=2^20. Because these numbers are relatively large, it may be a good idea to let N be the number whose binary representation is 100x, where x is a random 20-bit binary sequence generated by 20 flips of a fair virtual coin. Smaller numbers can be used, but be sure that N is always bigger than P[n] (by several orders of magnitude if possible). Use the relative frequency of winning (the number of times the game is won divided by the number of times the game is played) to estimate the probability of winning this game.

Compute the following theoretical estimate of winning this game, which may be derived as an advanced exercise, and compare to the estimate obtained by simulation.

Teacher Resource (Background Information)

Probability(win) = 1 - Probability(lose), and Probability(lose) is approximately

$$\left(1 - \frac{1}{2P[1]}\right) * \left(1 - \frac{1}{2P[2]}\right) * \ldots * \left(1 - \frac{1}{2P[n]}\right)$$

that is, the product of the n factors of the form one minus the reciprocal of twice P[i] for each i (1<= i <= n). Assume that each coin flip and possible divisibility check is an independent event. (This assumption is not quite true and is in fact the reason our formula is merely an estimate of the probability of losing, but it is also the reason that we can derive such a nice formula.) Then the probability of losing is the product of the probabilities of losing at each prime P[i]. There is a 1/2 chance that the result of the coin flip is tails, and if it is heads, which also occurs with a 1/2 chance, there is a 1/P[i] chance that N is divisible by P[i]. Hence, the probability of losing at a given prime P[i] is

$$\frac{1}{2} + \frac{1}{2} * \frac{P[i] - 1}{P[i]}$$

and a straightforward algebraic simplification gives proper form for the formula above.

Because we are taking a product of many numbers which are very close to 1, it is advisable to use logarithms in computing the estimate for the probability of losing to prevent underflow errors, that is, use the fact that the log of the product is the sum of the logs to find the log of the probability of losing, and then exponentiate to recover the probability of losing.

The prime number theorem states that the number of prime numbers in the interval (0,x] is approximately x/ln(x) for large x (ln is the natural logarithm function). Use the prime number theorem to estimate that probability of losing the game due to N being prime when N is in the interval [M-e,M+e]. As an option one can derive the following formula for this estimated probability. Probability(N in the interval [M-e,M+e] is prime) is approximately

$$\frac{M + e}{2e * \ln(M+e)} - \frac{M - e}{2e * \ln(M - e)}$$

which is derived by using the prime number theorem to estimate the number of primes in the intervals [0,M+e] and [0,M-e], subtracting the latter from the former to obtain an estimate of the number of primes in the interval [M-e,M+e], and dividing by the number of integers in, i.e., the length of the interval [M-e,M+e].

Note well that this estimates the probability of losing due to the particular circumstance that N is prime; there are other ways to lose when N is not prime, and hence, 1 minus this probability does not give the probability of winning. Still, this estimate does give an estimate of the best case scenario, that is, the probability of winning the game can be no larger than 1 - probability (N is prime).

Performance Assessment:

The students will submit a correct program, flowchart and discussion of the solution strategy implemented in their simulation algorithm.

Level 3: The presentation offers clear evidence of a thorough understanding of the mathematics related to the activity.

. Simulation program is correct.
. Discussion of the strategy for the solution algorithm is correct and consistent with the flowchart submitted.
. Correct and consistent symbolism is present in the flowchart submitted.

Level 2: The presentation offers evidence of substantial understanding of the mathematics related to the activity.

. Discussion of the strategy for the solution algorithm is correct but is inconsistent with the flowchart submitted.

or

. The flowchart submitted depicts a correct solution strategy but no complete, consistent and correct written discussion is submitted.

Level 1: The presentation offers a reasonable understanding of the necessary structure of a correct solution strategy.

Extension/Follow Up:

Use a weighted virtual coin or a virtual d-sided die instead of a fair coin to play the game.

Vary the parameters e and M for the range of N. Vary the set of primes used. Note that in all of these cases, the theoretical estimates will have slightly different formulas.

Invite a professional mathematician or computer scientist to make a related presentation .

Authors:

Carvel LaCurts
Pocomoke High School
Worcester County, Maryland